An embodied model of infant gaze-following

J. Gregory Trafton
(greg.trafton@nrl.navy.mil)
Naval Research Laboratory
Washington, DC 20375 USA

Anthony M. Harrison
(anthony.harrison@nrl.navy.mil)
Naval Research Laboratory
Washington, DC 20375 USA

Benjamin Fransen
(fransen@aic.nrl.navy.mil)
Naval Research Laboratory
Washington, DC 20375 USA

Magdalena Bugajska
(magda.bugajska @nrl.navy.mil)
Naval Research Laboratory
Washington, DC 20375 USA

Abstract

We present an embodied model of gaze-following. The model learns how to follow another's gaze by using cognitively plausible mechanisms. It matches a classic gaze-following experiment (Corkum & Moore, 1998) and runs on an embodied robotic system.

Keywords: infant gaze-following; embodied cognition; robotics; cognitive architectures

Introduction

Gaze-following is an important, early component of joint visual attention (Scaife & Bruner, 1975; Butterworth & Jarrett, 1991). Joint visual attention is looking at the same object as another person. Some researchers have suggested that joint visual attention is strongly related to the ability to infer others' mental states (Baron-Cohen, 1995). More recently, researchers have suggested that gaze following does not require a representational component (Woodward, 2003).

In fact, several researchers have recently built computational models to explore the emergence and learning of gaze-following.

Previous models of gaze-following

One of the challenges confronting models of gaze-following is to create an embodied model. Embodiment is important in this domain for a number of reasons. First, there has recently been a movement for embodied models of cognition (e.g., Wilson, 2000). Second, spatial and developmental models seem to be particularly amenable to embodied cognition. Third, embodied cognition forces an integrative approach across models, theories, and empirical results. Finally, the complexity of the physical world provides strong tests for the theory under question. Each of the models of gaze following (including ours) claims they have embodied characteristics. There are three existing models of the acquisition of gaze-following.

Nagai, Hosoda, Morita, & Asada (2003) used a neural network approach to learn that shifts in the caregiver's head pose pointed to a salient and interesting object. Over time, the model (which also runs on a robot) learned to follow the gaze of the caregiver to an interesting object.

Doniec, Sun, & Scassellati (2006) greatly sped up the algorithm by using pointing gestures to acquire joint attention. Their algorithm (which also ran on a robot) had the robot actively point to the object it thought the caregiver was gazing at. This pointing greatly increased learning rate through positive examples. The fact that infants start to make deictic gestures around 10 months of age (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979), which is about the same age that gaze-following is acquired (Corkum & Moore, 1995; Corkum & Moore, 1998) provides empirical evidence that infant gesture may be a component of gaze-following. Beyond this interesting suggestion, however, Doniec et al.'s primary contribution is that it is able to learn at a much faster rate than previous models.

Triesch, Teuscher, Deak, & Carlson (2006) also developed a model of gaze-following. Triesch et al.'s model monitors the caregiver's direction of gaze and gradually learns that the caregiver looks at objects in the environment that are interesting or novel to the infant, which is rewarding. Triesch et al. modeled the learning process through Temporal-Difference (TD) learning, a biologically plausible reinforcement learning algorithm. Triesch et al.'s model used a model of habituation to determine when to shift attention and learned to follow gaze to determine where optimal (most interesting) objects were in the environment. Their model used a simple grid world where objects could only exist in a limited number of locations.

It is a mantra in the modeling community that no model is perfect; future models attempt to improve upon past models. All three of these models made strong progress toward the understanding of gaze-following. Their biggest weakness, however, is that they had significant issues with cognitive plausibility. In order to show cognitive plausibility, we (1) use and integrate a variety of cognitively plausible mechanisms (e.g., models of human memory, attention, etc.), (2) run models using a similar experimental paradigm, and (3) match experimental data using those mechanisms within the constraints of the experimental paradigm.

Several criticisms have been leveled against the Nagai et al. model. First, that model required an extremely large amount of training data; probably too much to be cognitively plausible (Doniec et al., 2006). Second, their model does not seem to be able to scale up to the more representational stage of gaze-following (Butterworth & Jarrett, 1991). Third, their model seems to work for only a single caregiver (Doniec et al., 2006).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
An embodied model of infant gaze-following	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,4555 OVerlook Ave SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
(in press) 9th International Conference of Cognitive Modeling (ICCM 2009), 24-26 Jul, Manchester, UK

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

Doniec et al.'s model was built in a manner that did not emphasize cognitive plausibility; their focus was on achieving fast and efficient learning for gaze-following in a realistic embodied context. One aspect of their model that limits its plausibility as a cognitive developmental account is the fact that they used six objects (toys) for joint gaze-following. If we assume that their model is approximately a 10 m. old infant, it is well known that infants at that age can not reliably identify objects a caregiver is gazing at if there are other objects in the line of sight (Butterworth & Jarrett, 1991).

While we agree with many aspects of Triesch et al.'s model, several criticisms have also been leveled at it. Some researchers have explicitly questioned the psychological plausibility (Moore, 2006). Specifically, Moore suggested that accurately modeling the attentional processes of infants during gaze following is a critical component to psychological plausibility in gaze-following. Additionally, because Triesch et al. used a grid system to simplify the training, the need for spatial cognition was greatly reduced. Thus, according to critics, a more robust and/or psychological representation of space was needed (Doniec et al., 2006; Moore, 2006).

The goal of this project is to show how an embodied model of gaze-following can not only perform gaze-following but also have a higher degree of cognitive plausibility by having cognitive attentional mechanisms (Doniec et al., 2006; Moore, 2006), a spatial representation (Doniec et al., 2006; Moore, 2006), and a match to data. While a match to data is not a perfect measure of cognitive plausibility (Cassimatis, Bello, & Langley, 2008), it can be used to differentiate models. At the least, if a model can show performance and competence as well as a reasonable data fit, it is more plausible (and, to us, preferred), than a model that does not.

The data we attempt to match is an experiment by Corkum and Moore (1998).

Method (Corkum & Moore, 1998)

A complete description of the experiment can be found in Corkum & Moore (1998).

Participants

63 participants completed the study, 21 participants in each of three age groups (6—7, 8—9, and 10—11 month olds).

Setup and Procedure

The experiment took place in a cubicle where two toys had been placed. Each toy rested on a turntable on either side of the room. When activated, the toy lit up and the turntable rotated. Both toys were visible to the infant at all times.

At the beginning of the experiment, each child entered into the cubicle and sat on their parent's lap directly across from the experimenter. The experimenter sat .6 m away. The experimenter called the child's name or tickled the child's tummy to get the infant to look at the experimenter. After the child looked at the experimenter, the trial began.

Each trial consisted of the experimenter looking 90° left or right at one of the two toys. The experimenter gazed at the toy for 7 s. During the trial, the experimenter did not vocalize or touch the infant, nor did the experimenter call the infant's name.

The experiment consisted of three consecutive phases. In the baseline phase, there were four trials where the experimenter looked at a toy (two trials to each side). During the baseline phase the toy remained inactive (i.e., did not light up or turn) in order to assess spontaneous gaze-following.

During the shaping phase, there were four trials (two to each side), but this time, regardless of the infant's gaze, the toy that was gazed at by the experimenter lit up and rotated.

During the final testing phase, a maximum of 20 trials (10 to each side) occurred where the toy was activated only if the infant and the experimenter looked at the same toy. If the child successfully followed the experimenter's gaze 5 times in a row, the experiment terminated.

Scoring

Each head turn was coded as either a target (joint-gaze with the experimenter) or a non-target (the wrong toy was gazed at) response. Infant head turns that did not look at a toy (e.g., naval-gazing) were not scored.

Random gaze-following would correspond to approximately 50% accuracy. Accurate gaze-following would correspond to an accuracy rate significantly greater than 50%, while anti-gaze-following would correspond to an accuracy rate significantly less than 50%.

Results and Discussion

To maintain clarity and connection with other researchers who report accuracy, percentage scores will be reported here for both the baseline and the last four test trials instead of the reported difference scores.

As Figure 1 suggests, only 10—11 m infants could reliably follow gaze at baseline. After training, however, both 8—9 m and 10—11 m infants could reliably follow gaze (there was a slight, non-significant increase in gaze-following for the 6—7 m infants).

These results are consistent with other researchers (Corkum & Moore, 1995) who have shown that gaze-following reliably occurs during the end of the first year: only 10—11 m infants could reliably follow gaze at baseline. Interestingly, however, 8—9 m infants learned to follow gaze in the experimental setting with a modest amount of training.

Corkum and Moore (1998) interpret these data as showing that there are several precursors to gaze-following. First, infants must be mature enough to respond to different spatial locations; they must have some rudimentary spatial ability. Second, infants must be able to learn that an interesting event will occur where the person looks. They further suggest that the adult's head turn cues the infant's attention in the direction of the turn.

We next describe the architecture and the task model.

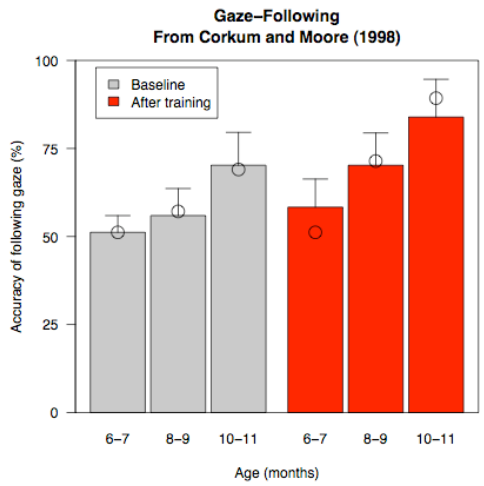

Figure 1: Experimental data from Corkum and Moore (1998). Bars are experimentall data and circles are model data. Error bars are 95% confidence intervals.

Architecture Description

ACT-R is a hybrid symbolic/sub-symbolic production-based system (Anderson, 2007). ACT-R consists of a number of modules, buffers, and a central pattern matcher. Modules contain a relatively specific cognitive faculty associated with a specific region of the brain. For each module, there are one or more buffers that communicate directly with that module as an interface to the rest of ACT-R. At any point in time, there may be at most one item in any individual buffer; thus, the module's job is to decide what and when to put a symbolic object into a buffer. The pattern matcher uses the contents of the buffer to match specific productions.

ACT-R supports the concept of purely bottom-up processing. Bottom-up or reactive processing occurs when there is no goal-directed processing that occurs. In contrast, top-down or goal-directed processing occurs when the goal buffer (intentional module) is part of the processing.

ACT-R interfaces with the outside world through the visual module, the aural module, the motor module, and the vocal module. Other current modules include the intentional, imaginal, temporal and declarative modules.

We have modified ACT-R by allowing it to perceive the physical world by attaching robotic sensors and effectors to it; we call our system ACT-R/E (the "E" is for Embodied). For ACT-R/E, we have added a new module (spatial) and modified the visual, aural and motor modules to work with our robot and to use real-world sensor modalities. We did not modify other parts of the architecture itself. Below we discuss the modifications to visual and motor (aural is not used in this project) and a brief description of the spatial module. Figure 2 shows a schematic of ACT-R/E.

Visual

The Visual Module is used to provide a model with information about what can be seen in the current environment. ACT-R normally sees information presented on a computer monitor. We modified the original visual module to accept input from a video camera. The visual module allows access to both the location of an object (the "where"' system) and a more detailed representation (the "what" system). Obtaining additional information about an object or person requires declarative retrieval(s). We used a 3D optical flow model to capture a person's 3D head pose in space and a fiducial tracker for object identification and localization. These systems are described more fully elsewhere (Kato, Billinghurst, Poupyrev, Imamoto, & Tachibana, 2000; Trafton, Bugajska, Fransen, & Ratwani, 2008; Fransen, Hebst, Harrison, & Trafton, under review).

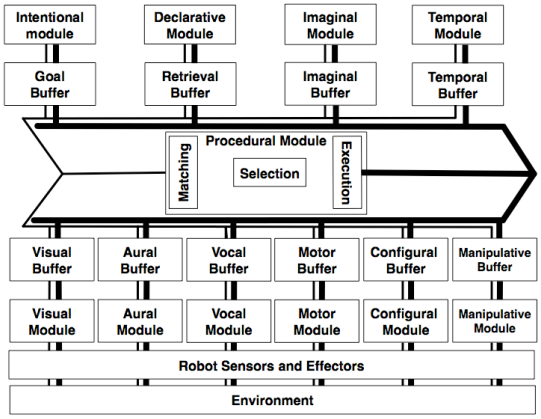

Figure 2: Schematic of ACT-R/E

Motor

Traditional ACT-R has a virtual motor system that allows virtual hand movements (e.g., typing, mouse movements). ACT-R/E's motor module allows commands to be issued for navigation and mobility, as well as providing self-localization knowledge. In this project, motor is used to control the robot's head, including the eyes and head pose.

Spatial

To facilitate acting in space, ACT-R/E utilizes a spatial theory called Specialized Egocentrically Coordinated Spaces (SECS, pronounced seeks) (Harrison & Schunn, 2003). SECS is neurologically inspired and based on 3D space (Previc, 1998). SECS provides two egocentric spatial modules, which are responsible for the encoding and transformation of representations in service of navigation (configural) and manipulation (manipulative).

The configural module provides high fidelity location information for attended representations that is automatically updated as the model moves through or looks around the environment. The configural module represents the world as spatial blobs that need to be navigated around, above, or below. These spatial blobs do not have a high degree of precision. The manipulative module uses a metric, geon-based 3D representation for objects. The manipulative module provides encodings of object geometry and orientation, a critical component to the gaze-following discussed below.

Simulator and Robot Description

Currently, the open-source Stage robot simulator (Collett, MacDonald, & Gerkey, 2005) is used to enable data collection and to speed-up the model development cycle.

Our current robot platform is the MDS (Mobile-Dexterous-Social) Robot (Breazeal, 2009). The MDS robot neck has 18 DoF for the neck and head including eye pitch and pan which allows the robot to look at various locations in 3D space. Perceptual inputs include a color video camera and a SR3000 camera to provide depth information. For the current project, the MDS head can move its eyes and head to look at various locations in 3D space.

Model Description

An ACT-R/E model was developed that simulates the development of gaze-following.

High Level Description of the gaze-following model

There are five model components that enable gaze-following: the reactive nature of the model; using ACT-R's memory system as a model of habituation; a more detailed description of the spatial components; the gaze-following itself; and the utility learning mechanism.

The reactive nature of the model The model itself is completely bottom-up; there is no goal-directed or top-down action in this model. The model was written in this manner because early gaze-following seems to be emergent rather than goal-directed (Triesch et al., 2006). Later models in the developmental process will need to have a goal-directed component.

Habituation in ACT-R When the model gazes at any object (person, toy, etc.), it looks at that object until it can recall the object before it attempts to look at a different object. This is an approximation of habituation (Sirois & Mareschal, 2002); several other researchers (Triesch et al., 2006) use an exponential function that is remarkably similar and formally equivalent to ACT-R's model of memory retrieval (Anderson, Bothell, Lebiere, & Matessa, 1998).

After the model gazes at and habituates to an object, it starts to look for a new object.

Spatial Module As mentioned earlier, standard ACT-R has only a rudimentary spatial ability. This ability is part of the visual module. In the visual module, a visual description of the object (a "what" component) and where that object is located in screen coordinates (a "where" component) is available (Byrne & Anderson, 1998). ACT-R's what and where system are used any time visual objects in the world need to be attended to. Many successful models of attention have been built using these mechanisms.

Unfortunately, the what and where components of ACT-R are not sufficient to follow gaze, much less provide even rudimentary spatial competency. As previously mentioned, two spatial modules were added to ACT-R, the configural module and the manipulative module.

The configural module is focused on the configuration of objects in the world relative to self. Specifically, it allows the model to determine how far away from self another object is and what angle that object is from self. Configural information changes dynamically as objects in the world change or move (including the self-model). This information is critical for navigation in general and spatial cognition in an embodied context.

For gaze-following, the manipulative buffer provides the orientation that a particular object is facing. Specifically, the manipulative buffer provides information about what direction a person is facing (body) or gazing (head).

The visual, configural, and manipulative modules are linked symbolically so that different types of spatial information about an object can be easily kept track of.

Gaze Following Gaze-following was implemented by adding constraints to the visual search mechanism. As implemented, gaze-following is a directed visual search along a retinotopic vector. Given a starting point and either an angle or an end point, the visual search will return the location on an object somewhere along that line within some tolerance. Note that this mechanism works in 3D space.

This simple mechanism allows the visual system to find candidate objects along a gaze, or any potential obstructions. These skills align nicely with Butterworth's developmental stages of gaze (Butterworth & Jarrett, 1991).

Utility Learning ACT-R is able to not only learn new facts and rules, but also to learn which rule should fire (called utility learning in ACT-R). It accomplishes this by learning which rule or set of rules lead to the highest reward. ACT-R uses an elaboration of the Rescorla-Wagner learning rule and the temporal-difference (TD) algorithm. The TD algorithm has been shown to be related to animal and human learning theory. The elaboration in ACT-R is more applicable for human learning and allows it to be more easily incorporated into a production-system framework (Fu & Anderson, 2006).

Briefly, any time a reward is given (e.g., for infants, a smile from a caregiver), a reward is propagated back in time through the rules that had an impact on the model getting that reward. Punishments may also be given with a similar time-course, but no punishments were given in this model.

For all models, we kept most of the ACT-R parameter defaults. The parameters that were changed include the base level learning (a decay value of .2 instead of the typical default of .5), which allowed for a reasonable habituation timecourse; utility noise (set at a reasonable .5) to allow low-use productions to occasionally fire; and the utility learning rate (set at .2) which allowed the productions to converge to a stable expected utility within a reasonable period of time (minutes instead of months).

A sample experimental model run

The first thing that the model does in an experimental trial is to find a person (called a caregiver in this example). This corresponds to the experimental procedure where the experimenter got the infant's attention (Corkum & Moore, 1998). The model looks at the caregiver until it has habituated to that person, as described above. The caregiver

looks at an object in the environment for 7 s or until the model makes a decision about where to look.

When the model is "young" it has a favored rule set, which is to locate, attend-to, and gaze at an object. The object can be anything in the model's field of view and it is chosen randomly.

If the caregiver is looking at the same object that the model decides to look at, the model is given a small reward. If the caregiver is looking at a different object than the model, no reward is given but the trial is completed and the reward process begins anew.

Even though there is a favored rule to find an object and gaze at it, the gaze-following rule competes with it. The gaze-following rule has a much lower utility when the model is young so it does not get an opportunity to fire very often. However, because of the relatively high noise value for utility (called expected-utility-noise in ACT-R), the gaze-following rule does occasionally get a chance to fire. If the gaze-following rule has a high enough utility to fire, it attempts to follow the gaze of the caregiver to an object.

The gaze-following production uses configural knowledge to determine the caregiver's distance and orientation from itself. As long as the model attends to the caregiver, the current information is available to the model.

The gaze-following production also uses manipulative knowledge of the head of the caregiver to determine what direction the caregiver's head is facing. This information is clearly important because without it the gaze of the caregiver could not be determined. Note also that the model assumes that the eyes are facing the same direction as the head. For the experimental procedure discussed here, this assumption is appropriate, but as children develop (by 1 year) they do differentiate between head pose and where the eyes themselves are gazing (Brooks & Meltzoff, 2002).

With this information, the infant model looks from the caregiver in the direction the head is facing. The model then finds the first available object in that direction, which is consistent with previous research (Butterworth & Jarrett, 1991). The model is again given a small reward. After habituation to that object, the trial ends and the model looks for another object to attend to.

Because the gaze-following production is correct more often than the random production (which is accurate on average 1/(number-of-objects), the gaze-following production slowly gains utility. However, it takes a period of time before the combination of noise and utility allow the gaze-following production to overtake and eventually become dominant over the random-object production.

Modeling developmental progress

When the model is young, it has a handful of productions that look around the world. Experience is simulated by concentrating gaze-following learning such that a few minutes is equal to 2 months. For the 6-7 m model, it was given 80 seconds of experience with looking around a simple world at objects and receiving feedback as described in the experimental run. For the 8-9 m model, three minutes of experience were given, and for the 10-11 six model, six minutes of experience were given. Because the rate of learning is dependent entirely on the utility learning rate parameter, learning occurred quite quickly in this model. Utility learning rate could be scaled down substantially to match actual infant learning time. In order to do this correctly, however, it would be important to know approximately how many times an infant attempts to follow a gaze or how often an infant receives feedback or the infant found something especially interesting to look at as well as knowledge about the environment (e.g., the number of objects). Other researchers have come to a similar conclusion concerning the importance of learning in gaze-following (Corkum & Moore, 1998; Triesch et al., 2006).

At each age (6-7, 8-9, and 10-11 m), the model was put through the exact same experimental procedure as Corkum & Moore (1998). Note that the lighting up and rotating of the toy provided a strong reward to the child, which is modeled by joint attention during the training phase of the procedure; no reward was given during the baseline phase, so this was a relatively pure measure of age-related ability.

To provide some match to the experimental procedure, 21 models (corresponding to the 21 participants) were run at each age group. However, to achieve stable results, the model was run 10 times with no utility learning for the baseline and after training conditions. This allowed the model to be tested after different age or experimental related amounts of practice yet maintain stable results.

Model fit

As is evident in Figure 1, the model matches the data quite well; $R^2 = .95$ and RMSD = .3. Critically, all model points are within 95% confidence intervals of the data. The model suggests that there is not a qualitative change in any child, but that as children gain more experience they get better at it. Interestingly, with a modest amount of experimental training, the 8-9 m model also showed improvement (though not, of course, as much as the 10-11 m model). Again the model suggests that the reason for this is that 8-9 m children were at the "right" developmental age to take advantage of the concentrated training. This training allowed productions that occasionally fired during "real life" to be focused and rewarded, which brought their utility to surpass the random behavior they had before the experiment started. Note again that the 6-7 m children did not statistically improve. The model explanation for this is that they simply had not had enough experience yet.

Embodied gaze following

The infant model at each stage of development was trained using Player and then run on an embodied platform (our robot). Movies are available at http://www.nrl.navy.mil/aic/iss/aas/CognitiveRobotsVideos.php.

General Discussion

We described an embodied model of gaze-following that is not only functional but matches data from a classic gaze-following paradigm and experiment. The primary advantage of this model over previous models is that it has a very high degree of cognitive plausibility. First, as Moore (2006) suggested, it has an accepted model of visual attention. Second, it has a psychologically plausible representation of space that is critical to the success of the model. Third, this model is embodied and runs on a physical robot, allowing additional tests of the theory as well as added complexity.

Of the model's 5 components (reactivity, habituation, the spatial module, gaze-following, and utility learning), three of them are absolutely critical to the success of the model. The reactivity nature of the module is a theoretical commitment to modeling young children, though the model could be written using a top-down model. Likewise, habituation is something that has been theoretically proposed and empirically observed, though it is not a critical component to the success of the model. The other three components, however, are needed. The spatial component integrates the spatial aspects of the task while the entire system could not function without the ability to perceive which direction a person is gazing. Because the developmental progress is accounted for by utility learning, it also is a necessary part of the model.

The model does make an interesting prediction: that 6 m infants (and even younger) could learn to follow gaze with enough practice. A core component to this prediction is that the infant have enough patience to go through enough training and the ability of young children to extract 3D information from the world. It is believed that 6 m olds do have this capability, but very young children do develop it.

This model also has several similarities to other infant data. The model does not understand obstructions and follows gaze to the first object along a path (Butterworth & Jarrett, 1991). The architecture does have the capability, however, to perform relatively precise gaze-following, ignoring highly salient objects in the path (the 'geometric' stage; Butterworth & Jarrett, 1991). The current model can not, however, follow gaze to a position outside its current field of view (the 'representational' stage). The current model has no true perspective-taking ability at all.

In order to provide the model with perspective taking abilities, it would presumably need more goal-directed cognition as well as more developed spatial capabilities.

Acknowledgments

This work was supported by the Office of Naval Research under funding document N0001409WX20173 to JGT.

References

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341-380.

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press, USA.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for. *Joint attention: Its origins and role in development*.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York, NY: Academic Press.

Breazeal, C. (2009). MDS Robot. http://robotic.media.mit.edu/projects/robots/mds/overview/overview.html.

Brooks, R., & Meltzoff, A. N. (2002). The Importance of Eyes: How Infants Interpret Adult Looking Behavior. *Developmental psychology*.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*.

Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson, & C. Lebiere (Eds.), *Atomic Components of thought* (pp. 167-200). Mahwah, NJ:: Lawrence Erlbaum.

Cassimatis, N. L., Bello, P., & Langley, P. (2008). Ability, Breadth and Parsmiony in Computational Models of Higher-Order Cognition. *Cognitive Science*, *32*(8), 1304-1322.

Collett, T. H., MacDonald, B. A., & Gerkey, B. P. (2005). Player 2.0: Toward a Practical Robot Programming Framework. In *Proceedings of the Australasian Conference on Robotics and Automation (ACRA 2005)*.

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, *34*(1), 28-38.

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In V. Corkum, & C. Moore (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83).

Doniec, M. W., Sun, G., & Scassellati, B. (2006). Active Learning of Joint Attention. *IEEE-RAS International Conference on Humanoid Robotics*

Fransen, B., Hebst, E., Harrison, A., & Trafton, J. G. (under review). 3D position and pose tracking. In.

Fu, W., & Anderson, J. (2006). From recurrent choice to skill learning: A model of reinforcement learning. *Journal of Experimental Psychology: General*.

Harrison, A. M., & Schunn, C. D. (2003). ACT-R/S: Look Ma, No "Cognitive-map"! In *Int'nl Conference on Cognitive Modeling*.

Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., & Tachibana, K. (2000). Virtual object manipulation on a table-top AR environment. In *IEEE and ACM International Symposium on Augmented Reality* (pp. 111-119).

Moore, C. (2006). Modeling the development of gaze following needs attention to space. *Developmental Science*.

Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003). A constructive model for the development of joint attention. *Connection Science*.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological Bulletin*, *124*(2), 123-164.

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*.

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, *6*(7), 293-298.

Trafton, J. G., Bugajska, M. D., Fransen, B. R., & Ratwani, R. M. (2008). Integrating vision and audition within a cognitive architecture to track conversations. *Proceedings of the Human Robot Interaction 2008 (HRI 2008)*.

Triesch, J., Teuscher, C., Deak, G. O., & Carlson, E. (2006). Gaze following: why (not) learn it? *Develop. Science*, *9*, 125-147.

Woodward, A. L. (2003). Infants'developing understanding of the link between looker and object. *Developmental Science*.